

FIG. 1

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3.5	0.121	0.036	0.467	0.664
4	0.096	0.06	0.519	0.708
4.5	0.078	0.079	0.565	0.745
5	0.062	0.098	0.615	0.782
5.5	0.05	0.127	0.659	0.813
6	0.04	0.163	0.694	0.836
6.5	0.033	0.202	0.725	0.855
7	0.025	0.248	0.763	0.878
7.5	0.021	0.304	0.78	0.889
8	0.015	0.368	0.816	0.909
8.5	0.012	0.418	0.836	0.92
9	0.009	0.512	0.856	0.93
9.5	0.007	0.581	0.863	0.934
10	0.006	0.679	0.835	0.919

FIG. 2

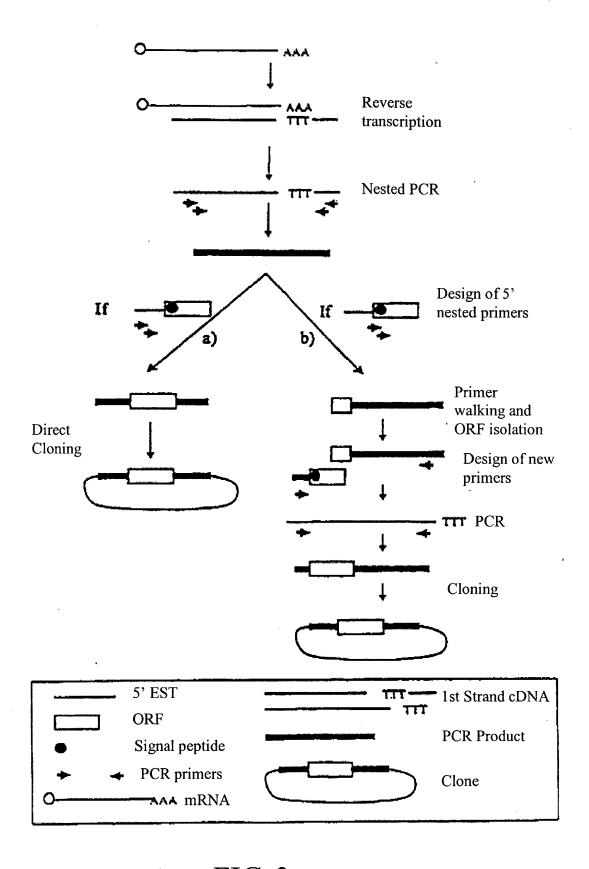
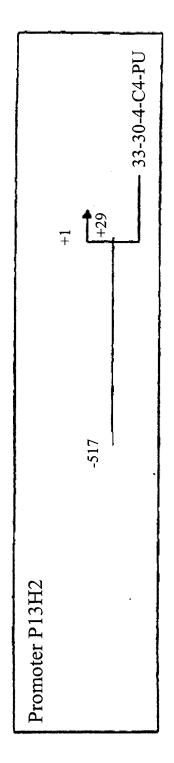
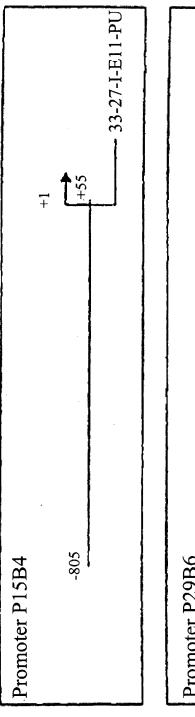


FIG. 3

Description of promoters structure isolated from SignalTag 5' ESTs





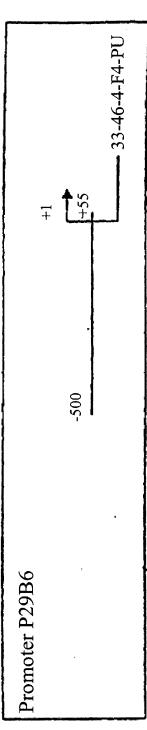


FIG. 4

## Description of Transcription Factor Binding Sites present on promoters isolated from SignalTag sequences

Promoter sequence P13H2 (546 bp):								
					•			
Matrix	Position	Orientation	Score	Length	Sequence	Location in:		
01 41 P. 01						SEQ ID NO: 17		
CMYB_01	-502	. <del>*</del>	0.983	9	TGTCAGTTG	17-25		
MYOD_Q6	-501	-	0.961	10	CCCAACTGAC	complement of 18-27		
S8_01	-444	• .	0.960	11	AATAGAATTAG	complement of 75-85		
S8_01	-425	+	0.966	11	AACTAAATTAG	94-104		
DELTAEF1_01	-390	•	0 960	11	GCACACCTCAG	complement of 129-139		
GATA_C	-364	•	0.964	11	AGATAAATCCA	complement of 155-165		
CMYB_01	-349	+	0.958	9	CTTCAGTTG	170-178		
GATA1_02	-343	+	0.959	14	TTGTAGATAGGACA	176-189		
GATA_C	-339	+	0.953	11	AGATAGGACAT	180-190		
TALIALPHAE47_01	-235	+ 1	0.973	16	CATAACAGATGGTAAG	284-299		
TALIBETAE47_01	-235	+	0.983	16	CATAACAGATGGTAAG	284-299		
TALIBETAITF2_01	-235	+	0.978	16	CATAACAGATGGTAAG	284-299		
MYOD_Q6	-232	•	0.954	10	ACCATCTGTT	complement of 287-296		
GATAI_04	-217	-	0.953	13	TCAAGATAAAGTA	complement of 302-314		
IK1_01	-126	+	0.963	13	AGTTGGGAATTCC	393-405		
IK2_01	-126	+	0.985	12	AGTTGGGAATTC	393-404		
CREL_01	-123	+	0.962	10	TGGGAATTCC	396-405		
GATAI 02	- <del>9</del> 6	+	0.950	14	TCAGTGATATGGCA	423-436		
SRY 02	-41	-	0.951	12	TAAAACAAAACA	complement of 478-489		
E2F 02	-33	+	0.957	8	TTTAGCGC	486-493		
MZF1_01	-5	-	0.975	8	TGAGGGGA	complement of \$14-521		
_								
Promoter sequence Pl						-		
Matrix	Position	Orientation	Score	Length	Sequence	Location in:		
						SEQ ID NO: 26		
NFY_Q6	-748	•	0.956	11	GGACCAATCAT	complement of 60-70		
MZF1_01	-738	+	0.962	8	CCTGGGGA	70-77		
CMYB_01	-684	+	0.994	9	TGACCGTTG	124-132		
VMYB_02	-682	•	0.985	9	TCCAACGGT	complement of 126-134		
STAT_01	-673	+	0.968	9	TTCCTGGAA	135-143		
STAT_01	-673	•	0.951	9	TTCCAGGAA	complement of 135-143		
MZF1_01	-556	•	0.956	8	TTGGGGGA	complement of 252-259		
IK2_01	-451	+	0.965	12	GAATGGGATTTC	357-368		
MZF1_01	-424	+	0.986	8	AGAGGGGA	384-391		
SRY_02	-398	-	0.955	12	GAAAACAAAACA	complement of 410-421		
MZF1_01	-216	+	0.960	8	GAAGGGGA	59 <b>2</b> -5 <b>9</b> 9		
MYOD_Q6	-190	+	0.981	10	AGCATCTGCC	618-627		
DELTAEFI_01	-176	. +	0.958	11	TCCCACCTTCC	632-642		
S8_01	. 5	-	0.992	11	C 4 C C C 4 4 TT 4 T			
MŽFI_01					GAGGCAATTAT	complement of 813-823		
	16	•	0.986	8	AGAGGGGA	complement of 813-823 complement of 824-831		
_		<u>.</u>						
Promoter sequence P2	9B6 (555 bp	,	0.986	8	AGAGGGGA	complement of 824-831		
_		- o): Orientation				complement of 824-831		
Promoter sequence P2 Matrix	9B6 (555 bp Position	Orientation	0.986 Score	8 Length	AGAGGGGA Sequence	Complement of 824-831  Location in: SEQ ID NO: 23		
Promoter sequence P2 Matrix ARNT_01	9B6 (555 bp Position -311	Orientation +	0.986 Score 0.964	Length	AGAGGGGA  Sequence GGACTCACGTGCTGCT	Location in: SEQ ID NO: 23 191-206		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01	9B6 (555 bp Position -311 -309	Orientation + +	0.986 Score 0.964 0.965	8 Length 16 12	AGAGGGGA  Sequence GGACTCACGTGCTGCT ACTCACGTGCTG	Location in: SEQ ID NO: 23 191-206 193-204		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01	29B6 (555 bp Position -311 -309 -309	Orientation + + + + + +	0.986 Score 0.964 0.965 0.985	8 Length 16 12 12	AGAGGGGA  Sequence  GGACTCACGTGCTGCT  ACTCACGTGCTG  ACTCACGTGCTG	Location in: SEQ ID NO: 23 191-206 193-204 193-204		
Promoter sequence P2 Matrix  ARNT_01  NMYC_01  USF_01  USF_01	9B6 (555 bp Position -311 -309 -309 -309	Orientation + + + -	0.986 Score 0.964 0.965 0.985 0.985	8 Length 16 12 12 12	AGAGGGGA  Sequence  GGACTCACGTGCTGCT  ACTCACGTGCTG  ACTCACGTGCTG  CAGCACGTGAGT	Complement of 824-831  Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 193-204		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01 USF_01 NMYC_01	.9B6 (555 bp Position -311 -309 -309 -309 -309	Orientation + +	0.986 Score 0.964 0.965 0.985 0.985	8 Length 16 12 12 12 12	AGAGGGGA  Sequence  GGACTCACGTGCTGCT ACTCACGTGCTG ACTCACGTGCTG CAGCACGTGAGT CAGCACGTGAGT	Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 193-204 complement of 193-204		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01 USF_01 NMYC_01 NMYC_01 MYCMAX_02	29B6 (555 bp Position -311 -309 -309 -309 -309 -309 -309	Orientation  + +	0.986 Score 0.964 0.965 0.985 0.985 0.956 0.972	8 Length 16 12 12 12 12 12	AGAGGGGA  Sequence  GGACTCACGTGCTGCT ACTCACGTGCTG ACTCACGTGCTG CAGCACGTGAGT CAGCACGTGAGT CAGCACGTGAGT CAGCACGTGAGT	Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 193-204 complement of 193-204 complement of 193-204		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01 USF_01 NMYC_01 NMYC_01 MYCMAX_02 USF_C	29B6 (555 bp Position -311 -309 -309 -309 -309 -309 -309 -309	Orientation + + +	0.986  Score  0.964 0.965 0.985 0.985 0.956 0.972 0.997	8 Length 16 12 12 12 12 12 12 8	AGAGGGGA  Sequence  GGACTCACGTGCTGCT ACTCACGTGCTG ACTCACGTGCTG CAGCACGTGAGT CAGCACGTGAGT CAGCACGTGAGT TCACGTGC	Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 193-204 complement of 193-204 complement of 193-204		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01 USF_01 NMYC_01 MYCMAX_02 USF_C USF_C	29B6 (555 bp Position -311 -309 -309 -309 -309 -309 -307 -307	Orientation  + +	0.986  Score  0.964 0.965 0.985 0.985 0.956 0.972 0.997 0.991	8 Length  16 12 12 12 12 12 12 8 8	AGAGGGGA  Sequence  GGACTCACGTGCTGCT ACTCACGTGCTG ACTCACGTGCTG CAGCACGTGAGT CAGCACGTGAGT CAGCACGTGAGT TCACGTGC GCACGTGA	Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 193-204		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01 USF_01 NMYC_01 MYCMAX_02 USF_C USF_C MZF1_01	-311 -309 -309 -309 -309 -309 -309 -307 -307 -292	Orientation + + +	0.986  Score  0.964 0.965 0.985 0.985 0.972 0.997 0.991 0.968	8 Length  16 12 12 12 12 12 8 8 8	AGAGGGGA  Sequence  GGACTCACGTGCTGCT ACTCACGTGCTG ACTCACGTGCTG CAGCACGTGAGT CAGCACGTGAGT CAGCACGTGAGT TCACGTGC GCACGTGA CATGGGGA	Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 193-204		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01 USF_01 NMYC_01 MYCMAX_02 USF_C USF_C USF_C MZF1_01 ELK1_02	-311 -309 -309 -309 -309 -309 -309 -307 -307 -292 -105	Orientation + + + + - + + -	0.986  Score  0.964 0.965 0.985 0.985 0.972 0.997 0.991 0.968 0.963	8 Length  16 12 12 12 12 12 12 8 8 8 14	AGAGGGGA  Sequence  GGACTCACGTGCTGCT ACTCACGTGCTG ACTCACGTGCTG CAGCACGTGAGT CAGCACGTGAGT CAGCACGTGAGT TCACGTGC GCACGTGA CATGGGGA CTCTCCGGAAGCCT	Complement of 824-831  Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 193-204 complement of 193-204 complement of 193-204 195-202 complement of 195-202 complement of 210-217 397-410		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01 USF_01 NMYC_01 MYCMAX_02 USF_C USF_C USF_C USF_C USF_C USF_C ELK1_01 ELK1_02 CETSIP54_01	-311 -309 -309 -309 -309 -309 -309 -307 -307 -292 -105 -102	Orientation + + +	0.986  Score  0.964 0.965 0.985 0.985 0.972 0.997 0.991 0.968 0.963 0.974	8 Length  16 12 12 12 12 12 12 14 10	AGAGGGGA  Sequence  GGACTCACGTGCTGCT ACTCACGTGCTG ACTCACGTGCTG CAGCACGTGAGT CAGCACGTGAGT CAGCACGTGAGT TCACGTGC GCACGTGA CATGGGGA CTCTCCGGAAGCCT TCCGGAAGCC	Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 195-202 complement of 210-217 397-410 400-409		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01 USF_01 NMYC_01 MYCMAX_02 USF_C USF_C MZF1_01 ELK1_02 CETS1P54_01 AP1_Q4	-311 -309 -309 -309 -309 -309 -309 -307 -307 -292 -105 -102 -42	Orientation + + + + - + + -	0.986  Score  0.964 0.965 0.985 0.985 0.972 0.997 0.991 0.968 0.963 0.974 0.963	8 Length  16 12 12 12 12 12 8 8 8 14 10 11	AGAGGGGA  Sequence  GGACTCACGTGCTGCT ACTCACGTGCTG ACTCACGTGCTG CAGCACGTGAGT CAGCACGTGAGT CAGCACGTGAGT TCACGTGC GCACGTGA CATGGGGA CATGGGGA CTCTCCGGAAGCCT TCCGGAAGCC AGTGACTGAAC	Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 193-204 complement of 193-204 complement of 193-204 complement of 193-202 complement of 210-217 397-410 400-409 complement of 460-470		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01 USF_01 NMYC_01 MYCMAX_02 USF_C USF_C USF_C MZF1_01 ELK1_02 CETS1P54_01 AP1_Q4 AP1FJ_Q2	-311 -309 -309 -309 -309 -309 -307 -307 -307 -292 -105 -102 -42	Orientation  + + + + +	0.986  Score  0.964 0.965 0.985 0.985 0.972 0.997 0.991 0.968 0.968 0.974 0.963 0.961	8 Length  16 12 12 12 12 12 8 8 8 14 10 11	AGAGGGGA  Sequence  GGACTCACGTGCTGCT ACTCACGTGCTG ACTCACGTGCTG CAGCACGTGAGT CAGCACGTGAGT CAGCACGTGAGT TCACGTGC GCACGTGA CATGGGGA CTCTCCGGAAGCCT TCCGGAAGCC AGTGACTGAAC AGTGACTGAAC	Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 193-204 complement of 193-204 complement of 193-204 complement of 195-202 complement of 210-217 397-410 400-409 complement of 460-470 complement of 460-470		
Promoter sequence P2 Matrix  ARNT_01 NMYC_01 USF_01 USF_01 NMYC_01 MYCMAX_02 USF_C USF_C MZF1_01 ELK1_02 CETS1P54_01 AP1_Q4	-311 -309 -309 -309 -309 -309 -309 -307 -307 -292 -105 -102 -42	Orientation + + + + +	0.986  Score  0.964 0.965 0.985 0.985 0.972 0.997 0.991 0.968 0.963 0.974 0.963	8 Length  16 12 12 12 12 12 8 8 8 14 10 11	AGAGGGGA  Sequence  GGACTCACGTGCTGCT ACTCACGTGCTG ACTCACGTGCTG CAGCACGTGAGT CAGCACGTGAGT CAGCACGTGAGT TCACGTGC GCACGTGA CATGGGGA CATGGGGA CTCTCCGGAAGCCT TCCGGAAGCC AGTGACTGAAC	Location in: SEQ ID NO: 23 191-206 193-204 193-204 complement of 193-204 complement of 193-204 complement of 193-204 complement of 193-202 complement of 210-217 397-410 400-409 complement of 460-470		